Web Intelligence in Higher Education: A Study on the usage of Business Intelligence techniques in Education

Adriano Maia, Filipe Portela and Manuel Filipe Santos
Algorithm Research Centre,
University of Minho,
Guimarães, Portugal
{cfp,mfs}@dsi.uminho.pt

***Abstract*— The usage of web intelligence systems in Education has been growing in the last years. The large amount of papers, dissertations and articles about this topic, created in the last decade, is a proof of the investment in the research of this theme. This paper has the main purpose of obtaining a better comprehension of the benefits and the disadvantages of implementing this type of systems to improve the way that we learn and teach. With this main goal in mind and considering that a simple analysis of a normalized database is not sufficient, a Web Intelligent System (WIS) was designed. This type of implementation allows the evaluation of large amounts of data (in this case related to Education), with a lot of incoherences, most of the time, which could be more easily analyzed with the help of the tools, supplied by these systems. This study was conducted with a sample of more than 133 students, that attended the course of Web Programming, in MiEGSI (master's degree in Information Systems) in University of Minho. With the WIS, it was possible to understand the main reasons that lead to success or retention in education, if this type of implementation was well accepted and used by the students, along with other important aspects, which can be verified by the results obtained. The application was accessed 21349 times, which means, in average, every student accessed it 143 times. The rate of retention is 10.5 % which corresponds to a total of 14 students that did not achieve success within the course. In average 25 students did not attend the classes, which links to the principal cause of retention.**

Keywords — Web Intelligence, ioEduc, Data Science, Big Data, Higher Education, Business Intelligence, Key Performance Indicators.

I. INTRODUCTION

In Portugal, the number of students in higher education has significantly increased, in the last few years [14]. This development could be explained by the increase of scholarships given, the growth and improvement of country's economic situation and the mentality of the general population, who want to enroll in higher education to be more qualified therefore improving their chances in the job market. However, in terms of grades and retention, the current situation and performance is still not at a desirable level. According to the data provided by the "Portuguese Republic" [14], in the year of 2017, retention has suffered a slight growth and the grades also recorded a drop, in terms of absolute value.

Since this type of problems has quite a few degrees of complexity tied to them, the solutions or proposals must be very carefully thought out and designed. The identification of the factors and measures to treat this low level of performance in the education field and the comprehension of how the relationship between student-teacher and the teaching process could be improved, are key points to improve student engagement. This type of difficulties/problems have been analyzed, in the last years, with the help of Business Intelligence (BI) techniques. This approach consists in transforming large amounts of data in relevant information to be used, posteriorly, in the decision-making process. For this to be achieved, a suitable, conceptual and digital solution to support all the variables is necessary. This work fits in a study about Big Data Analytics, and the solution proposed is a WIS. The data, presented and analyzed by this system, came from an application, used in the classroom by the students and teachers, called "ioEduc"[1]. The system includes data gathering, data analysis, results and interpretation. The development of this paper and the first phase of this study has the objective of understanding the main causes of retention and success. Furthermore, the objective will be to create and add, using data mining (DM), mechanisms to the "IoEduc" application, to predict and combat retention. This paper is divided into five sections: introduction; background and related work; web intelligence system; results; conclusions and final work.

II. BACKGROUND AND RELATED WORK

A. Web Intelligence

In the early 1990s and after the Internet and the World Wide Web (WWW) revolution, web-based information started to grow exponentially [8]. For Loh & Garin [4] this constant growth of web pages and web-based information lead to a phenomenon called information overload. In order to create

[1] "IoEduc" – Application developed by a startup called IOTECHPIS with the objective of helping students and teachers in the process of teaching and learning.

conditions for the business managers to make the right decisions and have a deeper understanding of the data, a new process, named Web Intelligence (WI), was designed to allow these business actors, using web-based tools and BI techniques, to analyze this data properly and achieve their goals [1]. Liu [3] states that WI connects the knowledge of two different environments, research and scientific development, with the objective of exploiting the major changes and improvements brought out by Artificial Intelligence (AI) in the next generation of web systems and services.

After some research we discover that there are already some WI solutions capable of being implemented and used in various areas:

- **OpenMIND** – open source tool that transforms data, from different sources in the web (also uses data from the Deep Web), in relevant data for decision making;
- **WebMIND** – powerful tool to search data within the web. This tool executes a process similar to the ETL process (Extract, Transform and Load) to enable data to be processed in other tools or be analyzed.
- **Verint WI platform** – these tools collect data from sources like blogs, websites, news reports websites, deep web and others, to discover relations between them.
- **Hiwire System** – in comparison with the other solutions presented, this one has the particularity of allowing the development to be made modularly. This type of construction facilitates the implementation in a varied set of areas.

B. Web Mining

Rashid Al-Azmi [8] defines Web Mining (WM) as an "automatic crawling and extraction of relevant information from the artefacts, activities and hidden patterns found in WWW". For Xu, Zhang & Li [13] WM corresponds to the act of using Data Mining (DM) techniques in the analysis of data stored in web repositories. They also add that this process could be divided into three categories:

- **Web content mining** – corresponds to the process of obtaining the most useful information from Web documents.
- **Web structure mining** – used to analyze the connection between web data (this data could be represented as web pages, web documents, web links, and others).
- **Web usage mining** – involves the application of DM techniques to understand interesting usage patterns from the WWW data and, with that, serve the needs, of the Web applications and users, better [7].

C. Web Usage Mining

Web Usage Mining (WUM) is an activity that focuses on discovering useful information in web-based data [7]. For Vellingiri, Kaliraj, Satheeshkumar, & Parthiban [11] WUM corresponds to the application of Data Mining (DM) techniques, in web data, to predict user behavior and obtain their main interests by analyzing the patterns. The sources of information, analyzed by WUM, could be obtained from data repositories like [7]: logs produced from web servers, logs produced from web proxy servers and logs produced from Browsers.

With the help of WUM we could obtain, through the characteristics of every project, some of the following information [11]:

- Number of hits;
- Number of visitors;
- Time and duration;
- Visitor referring website;
- Path analysis;
- Visitor IP address;
- Browser type;
- Cookies;
- Platform information.

WUM consists in four basic steps [7]:

- **Data Collection** – in this first step, the users' relevant data is collected from various sources;
- **Data Processing** – inconsistences, in the data collected, are treated through its nature;
- **Pattern Discovery** – application of techniques from DM, machine learning, statistics, and others, to find patterns;
- **Pattern Analysis** – this final step has the main purpose of finding knowledge from the patterns previously discovered. The irrelevant patterns are eliminated. This phase could use, for example, SQL (Structured Query Language) to find knowledge and to represent the results, for example, bar charts.

D. Student Retention

For Nelson, Clarke, Stoodley, & Creagh [5] the factors that lead or influence retention are complex and varied. The author also adds that success, retention and engagement are connected, therefore the reasons that lead to one could indicate or explain the causes of the other. Tinto [10], in is paper about student retention, suggests that universities need to recognize that the problems, demonstrated by students in the education environment, lie not only in them but also in the education actors (teachers, parents, and others). The comprehension of the causes and the connections between these three terms (student retention, student engagement and student retention), represents a step closer to reduce retention and bad behavior in education [12].

E. Student Success

Nelson et al. [5], using the study developed by Tinto [10], shows, with more detail than the four conditions (expectations, support, feedback and involvement), that there are five

categories to achieve success: learning, supporting, belonging, integrating and resourcing.

1. **Learning** – assessment, curricula, teaching practices and pedagogical styles;
2. **Supporting**- information, services, resources and peer support;
3. **Belonging** – interaction, inclusive activities and identity development
4. **Integrating**- academic literacies and personal literacies;
5. **Resourcing**- staff development, evidence base, communication and learning environments.

All these categories are measurable and with the right knowledge in analytics, this activity could smooth the comprehension of the most relevant causes of student success or retention [5].

F. Related Work

There are some studies addressing the usage of WI tools and BI techniques in Higher Education (HE). Guster & Brown [2] developed a BI system, that indicates to all users (students, teachers, and others), relevant metrics and strategic questions, that with the help of the results produced by the system, could be studied and analyzed (using dashboards),to improve the decision making process. Schierenbeck [9] shows us why the rise of BI and the usage of this technique could be a competitive weapon in HE environments. Furthermore, Piedade & Santos [6] developed a Student Relationship Management (SRM) system, based in a BI infrastructure. The system collects data about students and their academic behavior and is relevant for decision making associated to the teaching-learning process and in the automatic interaction with the students. Although the studies described before have the same base of study (academic education), to this project, the solution was not used in the classroom. They only use data provided from different academic sources. The data analyzed in this paper came from an application, called "IoEduc", used by students in one semester in "Web Programming" (academic course taught in University of Minho in the Information Systems master's degree). This study also provides insights about some of the most important factors of student retention/success.

III. WEB INTELLIGENCE SYSTEM

A. Methods and Material

In this study, the data generated by the usage of the application "IoEduc", in the first semester, by more than 133 students, was treated and analyzed. The database containing the data, is in the MySQL format and is called "webitclo_main". Since it is a database with data mainly from education usage, it focuses mostly in demographic characteristics (gender, name), educational characteristics (grades, name of discipline, retention, success, exams, questionnaires), application related factors (visits, date, validations, requisitions). Since the data is plentiful, only the data relevant to the kpis[2] defined was used. The tables used from the database were:

- **assessment_login** – data from the accesses, by group of work, to the application;
- **user_login** – general accesses to the application;
- **uc_presences** – data about attendance to the classes (theoretical or practical);
- **studentspwumtable** – information about the student;
- **1718_pwfinalresults** – information about the final results of each student;
- **mt_results** – grades of mini exams performed over the semester;

. To analyze the data, it was used Mysql Workbench and some Microsoft Tools and to represent them Tableau with Power Bi.

B. Methodology

For the development of the project as a whole, the Design Science Research (DSR) methodology was followed. To help the development of the WIS, Kimball's lifecycle methodology was used, to simplify and coordinate the work. This methodology consists on the creation of a set of activities, required to design, develop and deploy a data warehouse (DW) successfully [15]. For Kimball and Ross [15], and for the development of this project, this methodology could be divide in 8 activities: Planning; Define the business requirements; Design the multidimensional model; Design the technological architecture; Development and conception of the WIS; Implementation of the system; Test the system; Results. Nevertheless, this study also provides important information to the development of the WIS. The creation of this case study is inserted in "design and development phase" from DSR methodology.

C. Process

In the beginning of the work and after receiving the data and studying it, a set of indicators and kpis were defined, like, "number of students", "number of retentions", "causes of retention", and others that support all the development of the project. Based on what was said before, a WIS was designed and can be seen in the next figure.

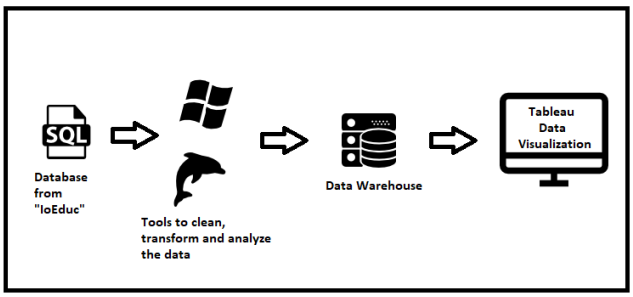

Figure 1 – Process to create a Web Intelligence System

This process is divided in three main aspects: data source, data preparation and data visualization. In the data source, the

[2] "Kpis" – Key Performance Indicators are measurable values used to demonstrate how effectively a company/organization is achieving their goals/key objectives.

data is received from the application “IoEduc”. In the data preparation, the information is first prepared to be analyzed (extracted, cleaned and transformed) and afterwards analyzed. In the last step, a set of dashboards and reports were provided, to enable a better and simpler analysis. The tables used from the database are described in section A. The Entity Relationship Diagram (ERD) is presented in the figure 2.

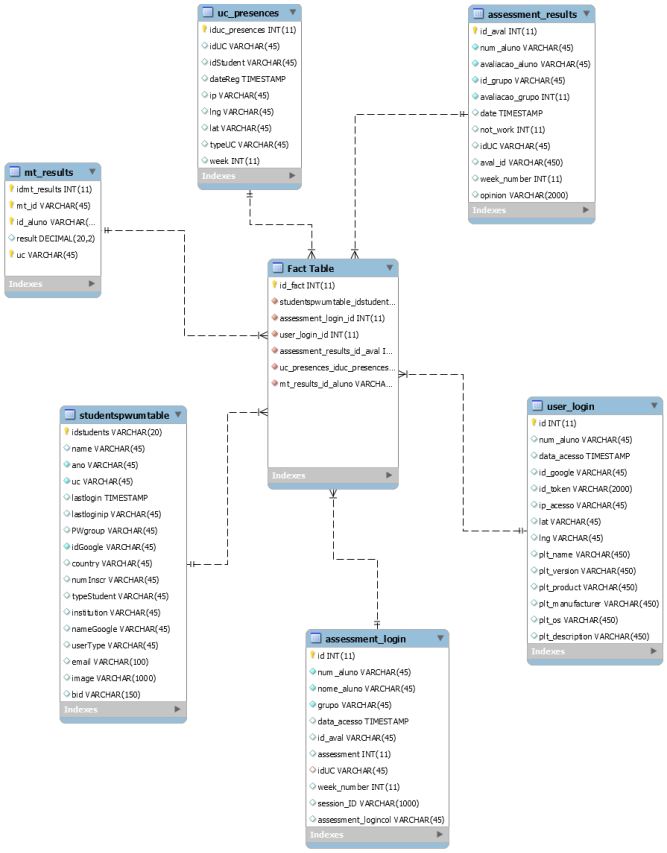

Figure 2 - Entity Relationship Diagram

IV. RESULTS

This study was conducted with data provide by an application (more properly a WIS) used in classroom, during the first semester of 2017/2018 in “Web Programming” (PW) (university course from Information Systems degree). In the following figure, a summary of the most important metrics concerning the access of the application “IoEduc” is presented. Analyzing the figure (3), we can observe that the total of users that accessed the application, since August of 2017 till February of 2018 (normally the duration of a semester) was 21349 and the number of distinct students, that access the application, was 149.

Figure 3 - Access metrics to the application "IoEduc" (1)

If we analyze the visits monthly, as showed in figure 4, we can see that the most visited month was October with 6280 visits. In the opposite pole is February with 62 visits, which is normal due to the fact of it being the last month of the semester and usually students only have one or two weeks of classes. It is also important to say that the semester began in September of 2017 and finished in February of 2018.

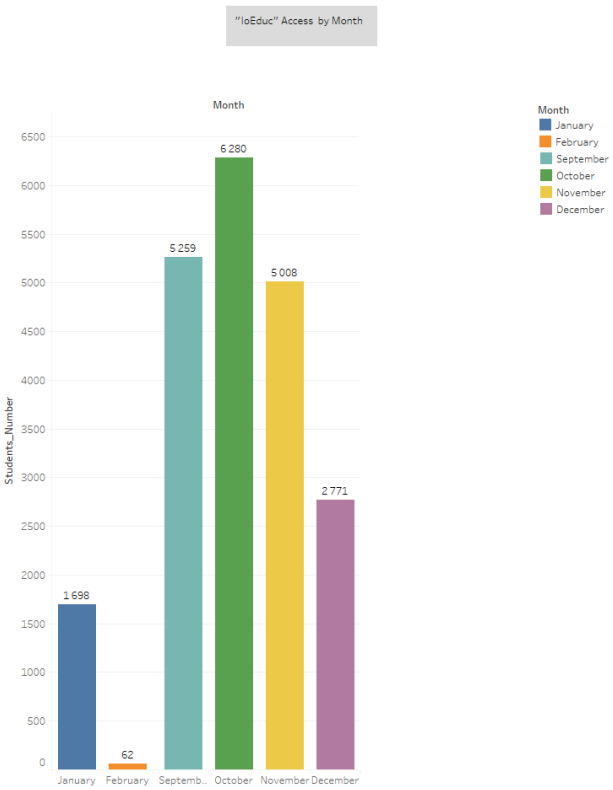

Figure 4 - Access metrics to the application "IoEduc" (2)

In figure 5 it is represented the type of the student: “Full time student” or “Part time student”. As we can see, most of the students (128), enrolled in PW are “Full time students”, and only 5 are “Part-time students”.

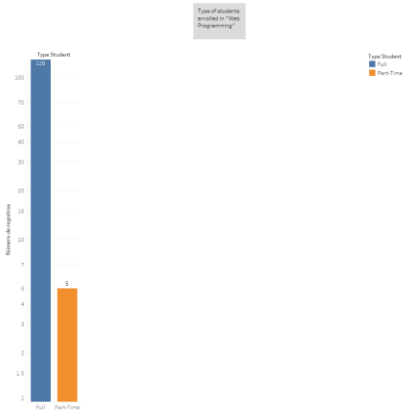

Figure 5 - Type of student

The figure 6 shows the distribution of the students by number of registrations in PW. The greatest slice (121) have only one

registration (which means that they are enrolled for the first time), 9 have two registrations and only 3 have three registrations, meaning that 12 students have not achieved success in previous enrollments to the discipline. It's important to say that, when a student fails the course, they are automatically enrolled in the next year.

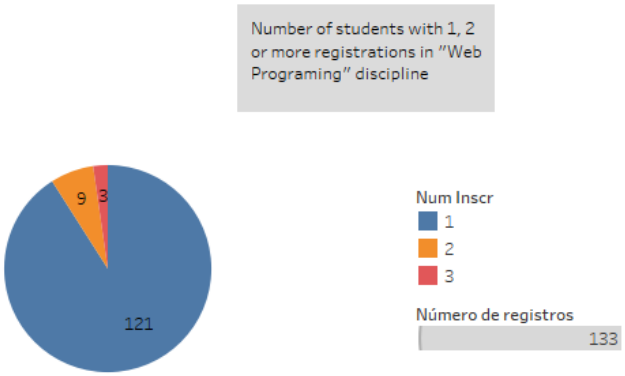

Figure 6 - Number of registrations in "Web Programming"

Lastly, in figure 7, we can visualize student presences by week of the year. Before analyzing this graph, we should consider the following points:

- This graph represents presences from theoretical classes;
- The first two weeks (31,36) collide with vacations, which explains the lower presences/access to the application;
- The weeks 37 and 41 have more presences because in those weeks there are more classes.

Considering what was said before, in average we have more than 25 absences (not counting the two first weeks). Also, if we not count the weeks that have more than one class, we can see that not in one week all the 133 students were present at the same time in class.

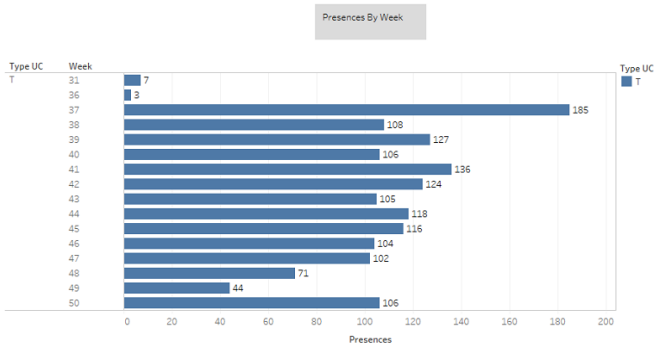

Figure 7 - Presences in classroom by week

Represented in figure 8 are the rates of success and retention. The rate of retention is 10.5 % and the rate of success was 89.5%, which means that in total 14 students didn't achieved the minimum grade (10 in a range of 0-20) to complete this discipline.

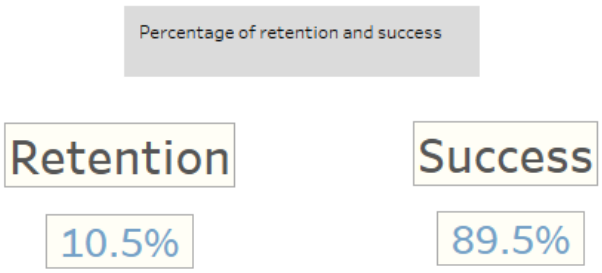

Figure 8 - Retention and success rates

Analyzing figure 9, we can see that the most represented grades are 13 and 15. A total of 55 students achieved a final grade of 17 or more points, which represents a rate of approximately 41%. A rate of 35% of the students had 13 or less values. The minimum grade (10) was obtained by 5 students and the maximum grade (20) was obtained by only 1 student. The column of the value "NA", represents the number of students that did not achieve success, which represents 14 students.

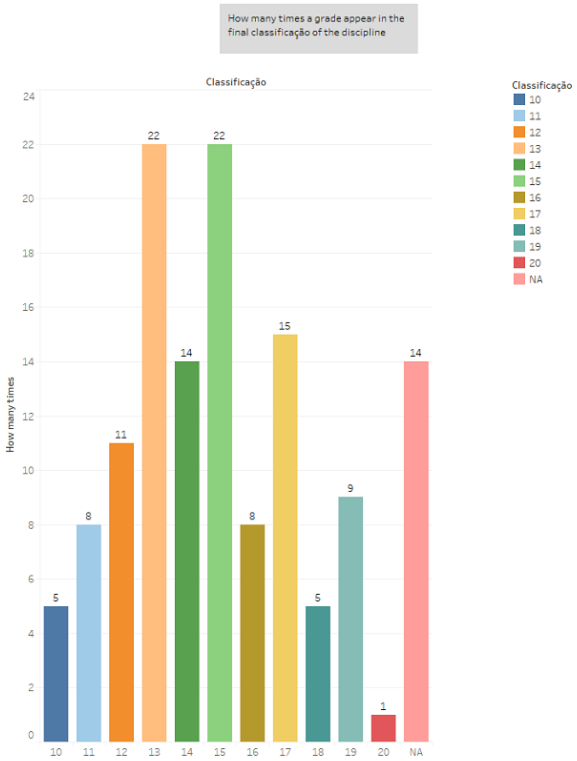

Figure 9 – Final Grades

In figure 10 its presented the number of attendance by student (both represented in the x line), week (represented in

the y line), grade and type of class (filters defined). In this analysis only, the students that did not achieve success (or final classification equals to "NA" see figure 9) are presented and the data related to theoretical classes. In a previous analysis it was said that 14 students didn't achieve success, but in this one only 11 are presented. This happens because 3 students didn't have any registration into the application throughout all the semester, making them automatically fail the course (due to excess of absences). Analyzing figure 10, we can see that 2 students have less than three registrations (in total) in the weeks under analysis. In a more embracing examination, almost every student has one or more absence in each week. In another way, only one student has at least one registration in every week. Furthermore, this could represent one of the most important factors of retention and explain the importance of being in every class to achieve success.

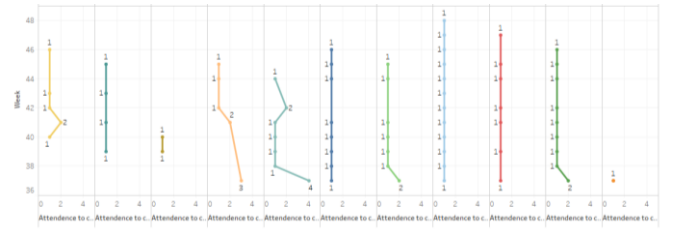

Figure 10 - Attendance in class by student, week and grade

V. Conclusions and Future Work

With the development of this work, it was possible to understand and analyze how Web Intelligence (WI) systems could influence Education, in terms of learning and teaching. At first, we could say that the implementation of this type of technology simplifies/ revolutionizes the way that absences are registered, the way material are provided, and the way exams are given. In terms of data produced, this type technology also ensures that the abundance and the quality of the data are guaranteed, which simplifies their analysis and with that facilitates the decision making process.

This study proved that students had a favorable reception to this new way of learning, based on the amount of data produced. In average each user accessed the application 143 times, with the most focus in the months of September to December (that matches with the time of the development of the group work and exams). In terms of absence of the classes, using the data provided and the total of students enrolled in the discipline, we reached the conclusion that, in average, 25 students are absent from the theoretical classes. These statistics could raise the question of success and retention, that represent 89.5% and 10.5% of the final grades, respectively. In regard to the students that achieved success, the most represented grades are 13 and 15 values. The maximum grade of 20 was only reached by 1 student.

Regarding future work, and since this was the first year of using a Web Intelligence (WI) system in the classroom, it is expected that the analysis of the results could continue in the next years, and with the implementation of more techniques, to analyze the data, like data mining, the process of decision making could become much simpler.

Acknowledgment

This work has been supported by COMPETE: POCI-01-0145-FEDER-007043 and FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2013. This work is also supported by the Deus ex Machina (DEM): Symbiotic technology for societal efficiency gains - NORTE-01-0145-FEDER-000026

References

[1] Brogden, J., Sinkwitz, H., Marks, D., & Orthous, G. (2014). Introduction to SAP BusinessObjects Web Intelligence 4.1 (3rd editio). Retrieved from https://s3-eu-west-1.amazonaws.com/gxmedia.galileo-press.de/leseproben/3673/Reading_Sample_sappress_1057_SAP_BusinessObjects_Web_Intelligence_updated.pdf

[2] Guster, D., & Brown, C. (2012). The Application of Business Intelligence to Higher Education: Technical and Managerial Perspectives. Journal of Information Technology Management, 23(1), 42–62. Retrieved from http://jitm.ubalt.edu/XXIII-2/article4.pdf

[3] Liu, J. (2003). Web Intelligence (WI): What makes wisdom web? IJCAI International Joint Conference on Artificial Intelligence.

[4] Loh, S., & Garin, R. (2001). WEB INTELLIGENCE – INTELIGÊNCIA ARTIFICIAL PARA DESCOBERTA DE CONHECIMENTO NA WEB. Intelligence. Retrieved from file:///C:/Users/adria/Google Drive/Tese/WI/5630eebc08ae13bc6c3549eb.pdf

[5] Nelson, K., Clarke, J., Stoodley, I., & Creagh, T. (2014). Establishing a framework for transforming student engagement, success and retention in higher education institutions. Findings Forum. Retrieved from http://studentengagementmaturitymodel.net/

[6] Piedade, M. B., & Santos, M. Y. (2009). Business Intelligence supporting the teaching-learning process. Proceedings of the 9th Wseas International Conference on Simulation Modelling and Optimization, 292–297.

[7] Rao, R. S., & Arora, J. (2017). A Survey on Methods used in Web Usage Mining. International Research Journal of Engineering and Technology (IRJET), 4(5), 2627–2631. https://doi.org/10.1109/ICITSI.2014.7048235

[8] Rashid Al-Azmi, A.-A. (2013). Data, Text and Web Mining for Business Intelligence: A Survey. International Journal of Data Mining & Knowledge Management Process (IJDKP), 3(2), 1–21. https://doi.org/10.5121/ijdkp.2013.3201

[9] Schierenbeck, C. (2013). Business Intelligence in Higher Education. In Fixing Higher Education (pp. 143–151). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-00213-8_11

[10] Tinto, V. (2005). Taking student success seriously: Rethinking the first year of college. Ninth Annual Intersession Academic Affairs Forum, 19, 1–8. https://doi.org/10.2307/1176972

[11] Vellingiri, J., Kaliraj, S., Satheeshkumar, S., & Parthiban, T. (2014). A Novel Approach for User Navigation Pattern Discovery and Analysis for Web Usage Mining. Retrieved from https://pdfs.semanticscholar.org/808f/278802b904bbcfee96eff9be1980cdcdb1de.pdf

[12] West, D., Huijser, H., Heath, D., Lizzio, A., Toohey, D., Miles, C., … Bronnimann, J. (2016). Higher education teachers' experiences with learning analytics in relation to student retention. Australasian Journal Of Educational Technology, 32(5), 48–60. https://doi.org/10.14742/ajet.3435

[13] Xu, G., Zhang, Y., & Li, L. (2010). Web Mining and Social Networking: Techniques and Applications. Springer. Retrieved from https://books.google.com/books?id=mXo9zKeYa6cC&pgis=1

[14] Statical Results can be consulted in the official web page of the portuguese republic: http://infocursos.mec.pt

[15] Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit, The Definitive Guide to Dimensional Modeling*. *Vasa*. Wiley. https://doi.org/10.1145/945721.945741